Selection enhances protein evolvability by increasing mutational robustness and foldability

Zheng, Jia ; Guo, Ning ; Wagner, Andreas

Abstract: Natural selection can promote or hinder a population's evolvability—the ability to evolve new and adaptive phenotypes—but the underlying mechanisms are poorly understood. To examine how the strength of selection affects evolvability, we subjected populations of yellow fluorescent protein to directed evolution under different selection regimes and then evolved them toward the new phenotype of green fluorescence. Populations under strong selection for the yellow phenotype evolved the green phenotype most rapidly. They did so by accumulating mutations that increase both robustness to mutations and foldability. Under weak selection, neofunctionalizing mutations rose to higher frequency at first, but more frequent deleterious mutations undermined their eventual success. Our experiments show how selection can enhance evolvability by enhancing robustness and create the conditions necessary for evolutionary success.

DOI: https://doi.org/10.1126/science.abb5962

Title: Selection enhances protein evolvability by increasing mutational robustness and foldability

Authors: Jia Zheng[1,2], Ning Guo[3] & Andreas Wagner[*1,2,4]

Affiliations:

[1]Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland.

[2]Swiss Institute of Bioinformatics, Quartier Sorge-Batiment Genopode, Lausanne, Switzerland.

[3] Zwirnereistrasse 11, Wallisellen, Zurich, Switzerland

[4] The Santa Fe Institute, Santa Fe, New Mexico, USA.

*Corresponding author. Email: andreas.wagner@ieu.uzh.ch.

Abstract:

Natural selection can promote or hinder a population's evolvability – the ability to evolve new and adaptive phenotypes – but the underlying mechanisms are poorly understood. To examine how the strength of selection affects evolvability, we subjected populations of yellow fluorescent protein to directed evolution under different selection regimes, and then evolved them towards the new phenotype of green fluorescence. Populations under strong selection for the yellow phenotype evolved the green phenotype most rapidly. They did so by accumulating mutations that increase both robustness to mutations and foldability. Under weak selection, neofunctionalizing mutations rose to higher frequency at first, but more frequent deleterious mutations undermined their eventual success. Our experiments show how selection can enhance evolvability by enhancing robustness and create the conditions necessary for evolutionary success.

One Sentence Summary: Natural selection drives the evolution of protein evolvability by enhancing robustness and foldability.

Main Text:

Natural selection drives adaptation, but we still know little about its role in changing the evolvability of a trait or organism (*1*, *2*). On the one hand, strong selection for an ancestral phenotype may enhance evolvability for derived phenotypes, because it may favor mutations that enhance not just fitness but also evolvability. The available evidence is limited and indirect (*3*–*5*). For example, a cytochrome P450 BM3 variant engineered for greater stability and fitness buffers the destabilizing effect of mutations that are neofunctionalizing, i.e., that convey new protein activities (*5*, *6*).

On the other hand, strong selection may impair evolvability because it purges weakly deleterious mutations that can convey new functions (*7*). Evolutionary theory holds that selection helps populations find peaks in adaptive landscapes of fitness, which can be visualized as topological maps of peaks and valleys. Weakly deleterious "stepping stone" mutations may help a population traverse the valleys that separate different fitness peaks. Such valleys, which are caused by epistatic interactions between different mutations, are abundant in the adaptive landscapes of evolving proteins (*8*–*11*). Weak selection that purges only the most detrimental mutations can aid such valley-crossing. Consistent with this view, when the enzyme β-lactamase TEM-1 is subject to "intense neutral drift" during experimental evolution – multiple rounds of mutation and selection to preserve its native phenotype of ampicillin resistance – the evolution of resistance to the newly introduced antibiotic cefotaxime is accelerated (*3*, *12*). More generally, experimental evolution shows that a population evolves a derived phenotype faster if it harbors genetic variation in the loci that affect the ancestral phenotype (*3*, *13*, *14*). Such standing genetic variation may even be adaptive if it has little effect on the ancestral phenotype (*15*). Therefore,

conditionally neutral or weakly deleterious mutations can accelerate the adaptive evolution of a derived phenotype (*15*). Here we performed experiments aimed to find out whether strong or weak selection more effectively enhances evolvability.

Results

Strong selection leads to greater evolvability than weak or no selection

We evolved yellow fluorescent protein [YFP, a variant of a jellyfish fluorescent protein (*16*)] in *E.coli* from an ancestral phenotype (yellow fluorescence) to a derived phenotype (green fluorescence) (Fig. 1A). The protein was engineered to be well expressed in *E.coli* (*16*) but it is not native to *E.coli*, which minimizes interference with the native *E.coli* proteome. Studying evolvability in a single protein can help us analyze the causes of evolvability in molecular detail. In addition, the fluorescent phenotype permits us to control the strength of selection with precision, because fluorescence-activated cell sorting (FACS) can screen individual cells for their phenotype.

We subjected each of four replicate populations of *E. coli* expressing yellow fluorescent proteins to four rounds ('generations') of directed evolution. In phase I of our experiment (Fig. 1A), we selected for yellow fluorescence through either strong selection (populations *S*, the top ~20% of fluorescing cells survive) or weak selection (populations *W*, cells that fluoresce above background survive) (*17*). We also subjected four replicate populations of yellow fluorescent proteins to four rounds of directed evolution without any selection for fluorescence (populations *N,* subject only to neutral drift). After phase I evolution, we initialized phase II evolution by subjecting the same populations to another four rounds of directed evolution under selection for

green fluorescence (Fig. 1A) (*17*). We used mutagenic PCR to generate 0.84 amino-acid changing mutations per YFP molecule per generation (tables S1 to S3).

During phase I evolution, yellow fluorescence in the unselected populations *N* dropped rapidly and decreased to <5% of the ancestral YFP's intensity, indicating an accumulation of deleterious mutations (Fig. 1B). For populations under weak selection, *W*, the intensity of yellow fluorescence decreased after the first generation of evolution but remained constant in the next three generations (Fig. 1B). This suggests a mutation-selection balance between the production of new deleterious mutations and selection against such mutations. In contrast, yellow fluorescence in populations under strong selection, *S,* significantly increased by 92.5% after four generations of evolution (one-sided *t* test, $P < 0.001$; Fig. 1B), indicating a likely spread of beneficial mutations.

We genotyped ~500 to 1000 protein variants per population and generation (table S4), which revealed that our evolving populations harbored different amounts of genetic variation. Specifically, during phase I populations *W* accumulated more amino acid-changing mutations and greater genetic diversity than populations *S* (figs. S1 and S2). Because greater genetic diversity may facilitate adaptive evolution, we hypothesized that populations *W* may have greater potential than populations *S* to evolve green fluorescence during phase II evolution. However, the opposite was the case. Populations *S* reached significantly higher green fluorescence than populations *W* during the first two generations of phase II evolution (one-sided Dunnett test with single-step adjustment, $P<0.05$; Fig. 1C). In addition, populations *S* evolved a green (512 nm) emission peak more rapidly than populations *W* (fig. S3). Analogous differences exist between populations *W* and unselected populations *N* (Fig. 1C, and figs. S1 to S3). In sum, selection on an

ancestral phenotype, and in particular strong selection, facilitates the evolution of a derived phenotype.

Strong selection on an ancestral phenotype leads to the most rapid fixation of neofunctionalizing mutations

To find out why strong selection causes greater evolvability than weak selection, we studied the dynamics of genetic polymorphisms in each replicate population during phase II (fig. S4). Two mutations (G66S and Y204C) swept through each replicate *S*, *W* and *N* population. Because of their ubiquity, we refer to these two mutations as universal mutations. In addition, another 20 mutations attained a frequency exceeding 30% in at least one replicate of populations *S*, *W* and *N*. To determine whether all these (2+20) mutations are adaptive for green fluorescence, we engineered each of them into the ancestral YFP and measured their effects on green fluorescence. Only the two universal mutations G66S and Y204C caused green shifts of the emission peaks (Fig. 2A). Individually, these mutations caused a ~9-fold and ~2-fold increase in green fluorescence (fig. S5), and together they shifted the emission peak from yellow (530 nm) to green (512 nm; Fig. 2A). Thus, G66S and Y204C are the only neofunctionalizing mutations.

To further characterize the role of the remaining mutations, we engineered each of these mutations into the background of G66S+Y204C, referred to as genotype U for 'universal', and measured the effect of these 20 genotypes on green fluorescence. Nine mutants significantly enhanced green fluorescence in the background of U (two-sided Dunnett test with single-step adjustment, $P<0.001$; fig. S5), but none of these changed the emission spectrum (Fig. 2B). These mutations might increase the amount of soluble and functional fluorescent protein, which may help explain why green fluorescence also increased, albeit very modestly, during selection for

yellow fluorescence in phase I (Fig. 1B).

Because only two mutations are responsible for the green shift in phase II, we suspected that the rapid spread of these two mutations resulted in faster adaptation of the strong selection (*S*) populations compared to the weak selection (*W*) and unselected (*N*) populations. Indeed, populations *S* displayed a higher frequency of the universal mutations G66S andY204C, and of the universal genotype U in at least last three of the four phase II generations (Fig. 2C).

Because the *S*, *W*, and *N* populations were subject to identical selection pressures during phase II, these faster sweeps likely originated in differences between populations at the end of phase I. We investigated if the reason was that the neofunctionalizing mutations had already attained a higher frequency in populations *S* at the end of phase I. However, this was not the case. Both universal mutations had lower frequency in populations *S* at the end of phase I evolution, and one of them (Y204C) had a significantly lower frequency in populations *S* than in populations *W* and *N* (one-sided Dunnett test with single-step adjustment, $P < 0.001$; fig. S6).

Strong selection leads to greater mutational robustness and higher foldability than weak selection

To resolve this apparent paradox – faster spreading of the neofunctionalizing mutations G66S and Y204C despite their lower initial frequency – it is relevant that our evolving sequences are likely to acquire one or more new mutations (with probability 0.53) in every generation (table S2). In addition, mutations that are different from the two neofunctionalizing mutation are expected to arise by chance alone 412-fold more often than these two mutations (table S2). This means that most variants containing neofunctionalizing mutations will also accumulate many

other mutations, most of which are near-neutral or deleterious (*18*, *19*). Increased robustness to such slightly deleterious mutations would increase the fitness of genotypes carrying neofunctionalizing mutations, and thus enable their spreading. We thus hypothesized that populations *S* had acquired genetic changes that cause greater robustness to deleterious mutations.

To validate this hypothesis, we mutagenized populations *S*, *W* and *N* at the end of phase I, and determined the residual fluorescence and the frequency of fluorescence-positive variants after mutagenesis. Populations *S* had indeed acquired greater mutational robustness, in both the ancestral yellow phenotype and the derived green phenotype. Specifically, populations *S* retained significantly higher yellow fluorescence intensity than populations *W* and *N* after mutagenesis (two-sided Dunnett test with single-step adjustment, $P < 0.05$ and 0.01 for comparing *S* with *W* and *N*; Fig. 3A). In addition, the post-mutagenesis frequency of yellow-fluorescence positive variants in *S* populations was 1.24-fold and 13.0-fold higher than in *W* and *N* populations (Fig. 3B). Moreover, the post-mutagenesis frequency of high-green variants, which have higher green fluorescence than the ancestral YFP, was 3.8-fold and 326.4-fold higher in *S* populations than in *W* and *N* populations, respectively (Fig. 3B).

Because most deleterious mutations reduce protein solubility by causing protein misfolding and instability (*20*, *21*), we suspected that populations *S* have evolved the ability to buffer such mutations by harboring protein variants with especially high foldability (folding efficiency) or stability (the free energy required to unfold a protein). Folding efficiency and stability can be jointly quantified by the concentrations of soluble protein in vivo (*22*). We thus measured and compared the amount of soluble fluorescent proteins in populations *S*, *W* and *N* at the end of

phase I (*17*). The amount of soluble proteins in populations *S* relative to that of ancestral YFP is 1.9-fold and 3.8-fold higher than that in populations *W* and *N* (Fig. 3C). By contrast, the fraction of insoluble protein in populations *S* is 2.8-fold and 2.5-fold lower than that in populations *W* and *N* (Fig. 3D and fig. S7), respectively.

We also analyzed the role of foldability more directly by quantifying the refolding yield of fluorescent proteins after unfolding (*17*). Indeed, a higher percentage of unfolded fluorescent proteins in populations *S* refolded than in populations *W*, *N* and in ancestral YFP during 24 hours of refolding (Fig. 3E and fig. S8A). Also, fewer than 50% of unfolded proteins refolded correctly at 25 °C within 24 h (Fig. 3E), demonstrating that solubility is likely limited by correct protein folding. In addition, the foldability of populations *S* did not decrease after random mutagenesis, whereas it decreased in populations *W* and *N* (fig. S8B). This observation suggests that increased foldability also increases mutational robustness.

Foldability cannot be completely disentangled from protein stability, because many mutations affect both (*20*, *23*). Stability can be estimated by monitoring a protein's structural integrity over time, and we measured for all our populations the residual yellow fluorescence after 12 h of incubation at 37 °C, the temperature at which we had conducted our experiments. All populations retained more than 95% of yellow fluorescence (fig. S8, C and D). Populations *S* showed higher stability than populations *W*, *N* and ancestral YFP only at higher, unphysiological temperatures above 65 °C (fig. S8, E and F). In sum, our experiments suggest that the higher robustness of populations *S* is primarily caused by higher foldability.

Foldability-improving mutations result in greater mutational robustness and higher evolvability

We next aimed to identify the genetic changes that increased foldability and mutational robustness by examining our sequence data. We focused on the 2 neofunctionalizing and 20 non-neofunctionalizing mutations that reached a frequency exceeding 30% at the end of phase II in populations *S*, *W* and *N* (fig. S4). The best candidates among them are four variants known to improve foldability (F47L, F65L, V164A and I172V) (*23–25*). All of these mutations and a fifth one (K102E) reached a higher frequency at the end of phase I in *S* populations relative to both *W* and *N* populations (fig. S9).

When we unfolded these mutants and refolded them (*17*), all five mutants yielded more correctly refolded protein than ancestral YFP (Fig. 4A), and three of them also refolded more rapidly than ancestral YFP (table S5). In addition, the mutations cause increased protein solubility (figs. S10A and S11). Notably, four of them also increased foldability (figs. S10B and table S5) and two of them significantly increased solubility in the background U (one-sided Dunnett test with single-step adjustment, $P<0.05$; figs. S10 C and S11), which folds with similarly low yield (~ 30%) as ancestral YFP (fig. S10D and table S5). The mutants also increased thermostability, albeit only at unphysiologically high temperatures (fig. S10, E to G, and table S6), indicating that they probably do not increase solubility by improving thermostability. In sum, all five key mutations improve foldability or protein solubility in the ancestral YFP background, and four of them do so in the green fluorescent protein background U.

To estimate the mutational robustness of specific mutants, we used PCR to introduce random mutations into ancestral YFP and into each of the 22 high-frequency YFP variants. We then

calculated the percentage of fluorescence retained relative to fluorescence before mutagenesis. After random mutagenesis, four out of five foldability-improving mutants retained greater yellow fluorescence (by 13.3 to 25.2%) than ancestral YFP (Fig. 4B), which indicates that they increased robustness. In one of the mutants (K102E), the increase in robustness was marginally significant (two-sided Dunnett test with single-step adjustment, P=0.064; Fig. 4B). After random mutagenesis, none of the other 17 mutations significantly retained greater yellow fluorescence intensity than ancestral YFP (two-sided Dunnett test with single-step adjustment, P>0.05; fig. S12). In addition, random mutations in all foldability-improving mutants created significantly more variants that remained yellow-fluorescence-positive than ancestral YFP (two-sided Dunnett test with single-step adjustment, P<0.05; Fig. 4C). Thus, foldability-improving mutations enhance the mutational robustness of the original yellow-fluorescence phenotype.

In addition, the five foldability-improving mutants also enhanced the robustness of the derived green fluorescence phenotype. First, after mutagenesis they retained significantly higher green fluorescence intensity (15.2 to 35.3% higher) than randomly mutated ancestral YFP (two-sided Dunnett test with single-step adjustment, P<0.05; Fig. 4B). Second, random mutations in the foldability-improving mutants created significantly more variants that remained green-fluorescence-positive (14.6-37.5% more) than ancestral YFP (two-sided Dunnett test with single-step adjustment, P<0.01; Fig. 4C). In addition, three foldability-improving mutations (F47L, F65L and I172V) significantly improved green fluorescence (by more than 27.0%) when combined with the two neofunctionalizing mutations (two-sided Dunnett test with single-step adjustment, P<0.001; fig. S5). These results suggest that the foldability-enhancing mutations can accelerate a selective sweep of neofunctionalizing mutants during phase II. In support of this observation, the foldability-improving mutations achieved higher frequency in *S* populations

than in *W* and *N* populations during each generation of evolution (fig. S13). Moreover, populations *S* retained a greater percentage of cells with higher green fluorescence than the ancestor during phase II evolution (Fig. 5A).

The advantage of the foldability-enhancing mutations could be mainly caused by their effect on robustness or by their effect on increasing fluorescence. We distinguished these two possibilities by measuring the effect of each mutation on fluorescence with or without mutagenesis. For example, the mutant F47L alone caused a 1.067-fold increase in green fluorescence-positive cells relative to ancestral YFP. After random mutagenesis of the mutant F47L, the increase in green fluorescence-positive cells relative to the mutagenized ancestral YFP was 1.390-fold. Thus, mutational robustness was responsible for most [84.2 % = 100×(1.390-1.067)/(1.390-1.0)] of the mutant's benefit for green fluorescence phenotype. This holds also for the other mutations, where we estimate that on average >75% of the fluorescence benefit comes from increased robustness (Fig. 5B). Thus, foldability-improving mutations likely promoted the spreading of neofunctionalizing mutations by enhancing robustness. This is consistent with the observation that the foldability-improving mutations did not greatly increase specific green and yellow fluorescence on their own (table S7). In the genetic background U, three of these mutations (F47L, F65L and I172V) also increased specific green fluorescence by more than 14% (table S7), and two of them (F47L and F65L) may have done so because of their spatial proximity to the chromophore (fig. S14). Such improvements in specific green fluorescence might additionally help promote the fixation of U.

Selection can eliminate deleterious mutations despite the presence of robustness-enhancing mutations

Although robustness-enhancing mutations augment the advantage of other beneficial mutations, their interactions with deleterious mutations are more complex. If they completely mask the effects of a deleterious mutation, this mutation may be preserved. However, if the mutation remains somewhat deleterious even in the presence of robustness-enhancing mutations, it may still be eliminated by selection, and at a rate that depends on the strength of selection. To test whether selection in *S* populations was sufficiently strong to eliminate deleterious mutations, we first used nonsense mutations, which produce truncated protein isoforms (*26*). Indeed, the frequency of nonsense mutations remained 2.1-fold lower in *S* than in *W* populations and 18.7-fold lower than in *N* populations at the end of phase I (Fig. 5C).

To complement this analysis, we also used the FoldX algorithm to predict destabilizing mutations(*17*). Such destabilizing mutations too became depleted in *S* populations (figs. S15 to S17). Specifically, highly destabilizing mutations had a consistently lower frequency in *S* populations than in *W* and *N* populations throughout phase I evolution (figs. S15A toS17). At the end of phase I, this frequency was 2.4-fold and 9.7-fold lower in *S* populations than in *W* and *N* populations, respectively (figs. S15A). Also, the protein variants of populations *S* were on average more stable than those of populations *W* and *N* during phase I evolution (fig. S15B).

Because computational stability predictions may be inaccurate when a protein harbors multiple amino acid changes (*27*), we also experimentally measured the stability of fluorescent proteins expressed in our populations. In these experiments, proteins from *S* populations showed greater

stability at 65 to 80 °C than those from *W* and *N* populations (fig. S8E). This prevalence of more stable proteins in *S* populations also persisted after mutagenesis (fig. S8F). In sum, selection can purge strongly deleterious mutations even in the presence of robustness-enhancing mutations.

Robustness-enhancing mutations help increase genetic diversity in populations subject to selection

Though not completely preventing the purging of deleterious mutations, robustness-enhancing mutations can still increase genetic diversity within a population by helping a population tolerate some deleterious mutations. Specifically, we observed that in both *S* and *W* (but not *N*) populations, proteins that carried robustness-enhancing mutations also harbored a significantly greater number of other amino acid changes at the end of phase I (one-sided t-tests, $P<0.05$; fig. S18).

To further examine whether robustness-enhancing mutations help populations S tolerate deleterious mutations during phase II evolution, we focused on the 22 mutations that reached a frequency exceeding 30% at the evolutionary end point. Among these mutations we identified four strongly deleterious mutations that significantly reduced the fluorescence of the universal green fluorescent genotype U by more than 20% (E18G, M79L, R110S and N145S; two-sided Dunnett test with single-step adjustment, $P<0.001$; fig. S5). Three of these deleterious mutations occurred in populations *S*, whereas only one of them occurred in any other population. Notably, more than 85% of the genetic variants that harbored one of these mutations also harbored at least one of the five foldability-improving mutations at the evolutionary end point (fig. S19). This suggests that robustness-enhancing mutations helped populations *S* tolerate strongly deleterious mutations. It is also consistent with the observation that populations *S* experienced greater

increases in both the number of amino-acid changes and in genetic diversity than populations *W* and *N* during phase II evolution (Fig. 6, A and 6B).

Discussion

Neutral or weakly deleterious mutations can play an important role in adaptive evolution (*13*, *28*), because they can convey new functions. They also help populations respond rapidly to environmental changes (*3*, *7*, *12*, *13*) and traverse fitness valleys created by epistatic (non-additive) mutational interactions (*29*). Weak purifying selection facilitates the accumulation of such mutations (*7*). If they were central to the evolution of a derived green fluorescence phenotype, our weak selection regime should have resulted in higher evolvability of this phenotype. However, we found that strong selection led to a more rapid evolution of the green fluorescence phenotype (Fig. 1B).

To understand why, consider that many proteins are marginally foldable (*30*) and marginally stable (*31*). Also, most mutations accumulated during evolution will further reduce protein foldability and stability (*20*, *32*), especially under weak selection (*12*, *33*, *34*). To function well, most proteins must first fold correctly (high foldability) and maintain structural integrity after folding (high stability) (*35*), resulting in selection on the evolution of foldability and stability. As a result, strong selection can not only stop the erosion of foldability and stability by purging deleterious mutations, it also favors mutations that enhance both properties. This ability is especially important during the evolution of new phenotypes, because neofunctionalizing mutations usually destabilize proteins and reduce folding efficiency (*5*, *34*, *36*). In our experiments, strong selection favored mutations that increase foldability and, to an even greater extent, mutational robustness. In doing so, selection enhanced the penetrance of beneficial

mutations (Fig. 5A) and accelerated selective sweeps (Fig. 2C). In addition, foldability-improving mutations may also promote the fixation of neofunctionalizing mutations by improving their specific protein activity during adaptive evolution (table S7). We anticipate that our observations apply to the evolution of most proteins in which foldability and stability are important.

Our experiments required high mutation rates so that we could observe adaptive evolution on a laboratory time scale. Comparable mutation rates have been observed in viruses (*37*, *38*) and microbes, especially those that are challenged to evolve rapidly by environmental stressors (*39*–*42*). Such high mutation rates are also required for the evolution of robustness as a direct response to mutation pressure (*3*, *4*, *43*). However, because our evolvability-enhancing mutations increase both robustness and foldability (Fig. 6C), their direct fitness benefit (fig. S20) would help them accumulate also under much lower mutation rates. An important question is whether the advantage of strong selection for evolvability also persists at low mutation rates.

Our work highlights the important role of standing variation for evolvability (*13*, *14*, *20*, *44*–*46*). It shows that natural selection can play a crucial and active role in creating standing variation that is both beneficial and enhances evolvability—for example by increasing robustness to deleterious mutations. This contrasts with some theoretical and experimental work, in which first-order selection for fitness conflicts with second-order selection for robustness (*47*, *48*). We predict that evolution can avoid this conflict when mutations with both fitness-enhancing and evolvability-enhancing roles exist.

Our work also suggests experimental designs that select for beneficial mutations without

depleting genetic variation before selecting for a new phenotype. Such experiments could be further enhanced by starting with protein variants engineered for high robustness. Furthermore, our observations extend beyond bioengineering. We expect that the evolutionary rescue of populations after environmental challenges, like climate change, may be easier in cases where a population's evolutionary potential has been previously enhanced by strong selection. Most generally, our observations suggest that natural selection can create favorable conditions for Darwinian evolution.

Materials and methods summary

Plasmids and Strains

We used the plasmid vector pBAD202/D-TOPO (K4202-01, Invitrogen) for cloning and expressing YFP alleles. This vector contains a kanamycin resistance marker and an arabinose-inducible *ara*BAD promoter. We used *E. coli* strain BW27783 (CGSC 12119) as a host to enable the homogeneous expression of the arabinose-inducible *ara*BAD promoter (*49*).

Creating mutant libraries

We used the same mutagenesis protocol during phase I and phase II of our experiments (*17*). Specifically, we performed mutagenic PCR (using the primers MutafpF and MutafpR, table S8) to randomly introduce mutations into the coding region of YFP, inserted the resulting mutant pool into the vector backbone by ligation, and then electroporated the ligation product into electrocompetent cells (*17*). After electroporation, we immediately added 1 ml of pre-warmed SOC medium, incubated the recovering cell culture for 1.5 h at 37 °C with shaking at 220 rpm in

a 50 ml tube, and used the recovered culture for further experiments.

To avoid the accumulation of mutations in the plasmid outside the YFP coding region, we inserted randomly mutated YFP genes into a fresh plasmid backbone in each generation. In addition, to avoid mutations that might accumulate in the *E.coli* genome, we transformed mutated YFP gene pools into fresh *E.coli* competent cells in each generation.

Directed evolution

We induced fluorescent protein expression in evolving populations at 37 °C for 12 h (with shaking at 220 rpm) by using 0.2% arabinose (*17*). We used an Aria III cell sorter (BD Biosciences) to sort cells at 4 °C in the fluorescein isothiocyanate (FITC) channel (λ_{ex} = 488 nm, λ_{em} = 530 ± 15 nm; phase I evolution) or in the AmCyan channel (λex = 405 nm, λem = 525 ± 25 nm; phase II evolution) according to the selection criteria described in Fig.1. In phase I evolution, we collected ~10^6 selected cells in ~1 ml of cold phosphate-buffered saline (PBS) buffer at 4 °C for each replicate population (*17*). In phase II evolution, we first selected ~5×10^4 cells with the top 1% of green fluorescence intensity, re-grew the sorted cells, and repeated the sorting process by again selecting ~10^4 cells in the top 1% of green fluorescence intensity (*17*). In each round, we regrew sorted cells, and isolated plasmids from sorted cells and used them as templates for the next mutation-selection cycle and for single-molecule real-time (SMRT) sequencing (*17*). To prevent cell proliferation or death, we placed selected cells on ice before the subsequent steps.

Engineering YFP variants

We used whole-plasmid PCR to engineer single mutants and some double mutants by designing primers that carry the corresponding mutations (table S9) (*17*). We also used whole-plasmid PCR to engineer single mutations into the genetic backgrounds of G66S, Y204C or U (G66T+Y204C) by using the mutants G66S, Y204C or U as templates (*17*). We used Gibson Assembly Master Mix (E2611, NEB) to engineer the double mutant U (G66T+Y204C) by using the primers G66Sf/Y204Cr and Y204Cr/G66Sf (table S9) (*17*).

Fluorescence assay using flow cytometry

We grew evolving populations or engineered variants in 200 µl or 2 ml of LB with 0.2% arabinose to induce the expression of YFP variants in evolving populations or engineered YFP variants (*17*). We added 40 µl of a culture to 160 µl of cold PBS buffer, transferred 20 µl of the resulting suspension to 180 µl of cold PBS buffer, and mixed the solution thoroughly. We used the resulting mixture to measure yellow fluorescence (λ_{ex} = 488 nm and λ_{em} = 530 ± 15 nm) and green fluorescence (λ_{ex} = 405 nm and λ_{em} = 525 ± 25 nm) using flow cytometry. We performed fluorescence assays at room temperature with a flow rate of ~3000 events/s by using a Fortessa cell analyzer (BD Biosciences). We performed at least three biological replicates for each replicate population or each variant, and analyzed ~10^4 events per replicate. To prevent cell proliferation or death, we placed all samples on ice until we had finished all assays.

Flow cytometry data analysis

We performed flow cytometry data analysis by using FlowJo V10.4.2 (LLC). Briefly, we used forward scatter height (FSC-H) versus side scatter height (SSC-H) density plots to select a homogeneous cell population (p1 in fig. S21A), and used side scatter area (SSC-A) versus side

scatter height (SSC-H) density plots to exclude doublets (p2 in fig. S21B). We used the resulting filtered data p2 for determining the fluorescence intensity of evolving populations, and for determining mutational robustness of evolving populations and those single mutants (Figs. 1, 3, 4, and 5 and fig. S12). We used FITC-height versus AmCyan-height density plots to select the dominant cell population (p3 in fig. S21C), and used the resulting filtered data p3 for determining the fluorescence intensity of each mutant (figs. S5 and S20).

Protein solubility determination and refolding kinetics measurements

We induced the expression of YFP variants in evolving populations or engineered mutants by growing cells in 2 ml LB medium with 50 μg/ml of kanamycin and 0.2% arabinose in a 10 ml tube at 37 °C and at 220 rpm for 12 h (*17*). We used CelLytic™ B Cell Lysis Reagent (B7435-500ml, Sigma) to extract both soluble and insoluble proteins from the collected cells by following the manufacturer's protocol (*17*). We quantified the amount of both soluble and insoluble proteins in each sample by SDS–polyacrylamide gel electrophoresis (SDS-PAGE) (*17*).

To unfold proteins, we diluted 5 μl of crude lysate with 45 μl of 9M urea (containing 10 mM dithiothreitol) and incubated the solution at 95 °C for 5 min. To refold unfolded proteins, we diluted 10 μl of the unfolded samples with 180 μl of TNG buffer in a 96 well microplate, and used an infinite F200 Pro (λex = 485 nm, λem = 530 nm) or a Spark 10M (λex = 485 nm, λem = 530 nm and λex = 405 nm, λem = 512 nm) microplate reader to measure fluorescence intensity at ~20min intervals (*17*).

SMRT sequencing and sequence data analysis

We barcoded YFP variants of each replicate population by using PCR and ligation for SMRT sequencing and used the Pacific Biosciences RS2 instrument (Pacific Biosciences) to perform sequencing (*17*). We used the SMRTAnalysis v2.3 package (*50*) to perform primary data analysis (*17*). We wrote Python scripts (Python 2.7.12) to identify point mutations and their combinations, and to calculate genetic diversity in each replicate population from each generation of evolution (*17*).

Statistical analysis

Unless specified otherwise, we conducted pairwise comparisons by using a one-tailed t-test, and conducted multiple comparisons to a control by using a one-sided Dunnett test with single-step adjustment. We performed all statistical analysis using R version 3.4.1.

References and Notes:

1. M. Pigliucci, *Nat. Rev. Genet.* **9**, 75–82 (2008).
2. J. L. Payne, A. Wagner, *Nat. Rev. Genet.* **20**, 24–38 (2018).
3. S. Bershtein, K. Goldin, D. S. Tawfik, *J. Mol. Biol.* **379**, 1029–1044 (2008).
4. J. D. Bloom *et al.*, *BMC Biol.* **5**, 29 (2007).
5. J. D. Bloom, S. T. Labthavikul, C. R. Otey, F. H. Arnold, *Proc. Natl. Acad. Sci. U. S. A.* **103**, 5869–5874 (2006).
6. M. Soskine, D. S. Tawfik, *Nat. Rev. Genet.* **11**, 572–582 (2010).
7. M. A. Stiffler, D. R. Hekstra, R. Ranganathan, *Cell*. **160**, 882–892 (2015).
8. C. Bank, S. Matuszewski, R. T. Hietpas, J. D. Jensen, *Proc. Natl. Acad. Sci. U. S. A.* **113**, 14085–14090 (2016).

9. T. N. Starr, L. K. Picton, J. W. Thornton, *Nature*. **549**, 409–413 (2017).

10. J. Domingo, P. Baeza-Centurion, B. Lehner, *Annu. Rev. Genomics Hum. Genet.* **20**, 433–460 (2019).

11. D. M. Weinreich, N. F. Delaney, M. A. Depristo, D. L. Hartl, *Science.* **312**, 111–114 (2006).

12. S. Bershtein, D. S. Tawfik, *Mol. Biol. Evol.* **25**, 2311–2318 (2008).

13. E. J. Hayden, E. Ferrada, A. Wagner, *Nature*. **474**, 92–95 (2011).

14. J. Zheng, J. L. Payne, A. Wagner, *Science.* **365**, 347–353 (2019).

15. A. B. Paaby, M. V. Rockman, *Nat. Rev. Genet.* **15**, 247–258 (2014).

16. C. Stanton *et al.*, *Nat Chem Biol*. **10**, 99–105 (2014).

17. *See supplemental materials.*

18. K. S. Sarkisyan *et al.*, *Nature*. **533**, 397–401 (2016).

19. L. Rockah-Shmuel, Á. Tóth-Petróczy, D. S. Tawfik, *PLOS Comput. Biol.* **11**, e1004421 (2015).

20. N. Tokuriki, D. S. Tawfik, *Nature*. **459**, 668–673 (2009).

21. A. Bandyopadhyay *et al.*, *Nat. Chem. Biol.* **8**, 238–245 (2012).

22. R. G. Smock, I. Yadid, O. Dym, J. Clarke, D. S. Tawfik, *Cell*. **164**, 476–486 (2016).

23. J.-D. Pédelacq, S. Cabantous, T. Tran, T. C. Terwilliger, G. S. Waldo, *Nat. Biotechnol.* **24**, 79–88 (2006).

24. T. Hyeon Yoo, A. James Link, D. A. Tirrell, *Proc. Natl. Acad. Sci. U. S. A.* **104**, 13887–13890 (2007).

25. A. Miyawaki, T. Nagai, H. Mizuno, *Curr. Opin. Chem. Biol.* **7**, 557–562 (2003).

26. A. A. Vakhrusheva, M. D. Kazanov, A. A. Mironov, G. A. Bazykin, *J. Mol. Evol.* **72**,

138–146 (2011).

27. J. Yang *et al.*, *PLoS One*. **15**, e0233509 (2020).

28. E. Rigato, G. Fusco, *J. Exp. Zool. Part B Mol. Dev. Evol.* **326**, 31–37 (2016).

29. B. Steinberg, M. Ostermeier, *Sci. Adv.* **2**, e1500921 (2016).

30. S. Govindarajan, R. A. Goldstein, *Biopolymers*. **42**, 427–438 (1997).

31. D. M. Taverna, R. A. Goldstein, *Proteins*. **46**, 105–109 (2002).

32. R. Geller, S. Pechmann, A. Acevedo, R. Andino, J. Frydman, *Nat. Commun.* **9**, 1781 (2018).

33. S. Bershtein, M. Segal, R. Bekerman, N. Tokuriki, D. S. Tawfik, *Nature*. **444**, 929–932 (2006).

34. N. Tokuriki, D. S. Tawfik, *Curr. Opin. Struct. Biol.* **19**, 596–604 (2009).

35. J. M. Sanchez-Ruiz, *Biophys. Chem.* **148**, 1–15 (2010).

36. K. L. Petrie *et al.*, *Science*. **359**, 1542–1545 (2018).

37. A. S. Lauring, J. Frydman, R. Andino, *Nat. Rev. Microbiol.* **11**, 327–336 (2013).

38. S. Duffy, L. A. Shackelton, E. C. Holmes, *Nat. Rev. Genet.* **9**, 267–276 (2008).

39. F. Taddei *et al.*, *Nature*. **387**, 700–702 (1997).

40. J. Arjan G. *et al.*, *Science*. **283**, 404–406 (1999).

41. I. Bjedov *et al.*, *Science*. **300**, 1404–1409 (2003).

42. R. C. MacLean, C. Torres-Barceló, R. Moxon, *Nat. Rev. Genet.* **14**, 221–227 (2013).

43. E. van Nimwegen, J. P. Crutchfield, M. Huynen, *Proc. Natl. Acad. Sci. U. S. A.* **96**, 9716–9720 (1999).

44. R. D. Gupta, D. S. Tawfik, *Nat. Methods*. **5**, 939–942 (2008).

45. S. L. Rutherford, S. Lindquist, *Nature*. **396**, 336–342 (1998).

46. N. Rohner *et al.*, *Science.* **342**, 1372–1375 (2013).

47. C. O. Wilke, J. L. Wang, C. Ofria, R. E. Lenski, C. Adami, *Nature*. **412**, 331–333 (2001).

48. M. S. Johnson, A. Martsul, S. Kryazhimskiy, M. M. Desai, *Science.* **366**, 490–493 (2019).

49. A. Khlebnikov, K. A. Datsenko, T. Skaug, B. L. Wanner, J. D. Keasling, *Microbiology*. **147**, 3241–3247 (2001).

50. *PACBIO, Analytical solutions for PACBIO sequencing data. Pacific Biosciences DevNet; www.pacb.com/support/documentation//DevNet.*

51. N. Guo, J. Zheng, Selection_strength_evolvability. *Zenodo* (2020), https://doi.org/10.5281/zenodo.3819561.

52. D. J. Warren, *Anal. Biochem.* **413**, 206–207 (2011).

53. J. Schindelin *et al.*, *Nat. Methods*. **9**, 676–682 (2012).

54. S. Bratulic, F. Gerber, A. Wagner, *Proc. Natl. Acad. Sci. U. S. A.* **112**, 12758–63 (2015).

55. M. J. Chaisson, G. Tesler, *BMC Bioinformatics*. **13**, 238 (2012).

56. H. Li *et al.*, *Bioinformatics*. **25**, 2078–2079 (2009).

57. K. E. Kim *et al.*, *Sci. Data 1*, 140045 (2014).

58. W. Shao *et al.*, *J. Virol. Methods*. **203**, 73–80 (2014).

59. N. Tokuriki, F. Stricher, L. Serrano, D. S. Tawfik, *PLoS Comput. Biol.* **4**, e1000002 (2008).

60. A. Waterhouse *et al.*, *Nucleic Acids Res.* **46**, W296–W303 (2018).

61. E. De Meulenaere *et al.*, *J. Am. Chem. Soc.* **135**, 4061–4069 (2013).

62. B. G. Reid, G. C. Flynn, *Biochemistry*. **36**, 6786–6791 (1997).

Acknowledgments: We acknowledge the experimental support of the flow cytometry facility and the functional genomics center at the University of Zurich. We thank Dr. H. E. L. Lischer, Dr. C. Bello and Dr. G. Schweizer for the help on SMRT sequencing data analysis or protein stability analysis. We thank Prof. Y. Schaerli, J. Duarte, Dr. M. Olombrada Sacristan, Dr. B. Ravi, Dr. P. Dasmeh and K. Ramanadane for experimental assistance or comments on the manuscript. We thank Prof. B. Schuler and Prof. J. L. Payne for helpful discussions. **Funding:** This project has received funding from the European Research Council under grant agreement 739874, and from Swiss National Science Foundation grant 31003A_172887. **Author contributions:** J.Z. and A.W. designed the experiments. J.Z. performed the experiments. J.Z., N.G. and A.W. contributed to data analysis. J.Z. and A.W. wrote the paper. All authors read and edited the paper. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or supplementary materials. SMRT sequencing data are available at DDBJ, EMBL, and GenBank under accession numbers KEEO00000000 and KCZY00000000. Custom code used in this study is available at Zenodo under the accession number 3819561 (*51*).

Legends

Fig. 1. Experimental evolution of yellow fluorescent protein. (**A)** We subjected four replicate *E. coli* populations for each experimental treatment to directed evolution under selection for yellow fluorescence (phase I, λex = 488 nm and λem = 530 ± 15 nm). After four mutation-selection cycles, we continued directed evolution for four more cycles but under selection for green fluorescence (phase II, λex = 405 nm and λem = 525 ± 25 nm) by selecting the top 0.01% of cells in each generation. The areas shaded in yellow or green indicate the proportion of a population allowed to survive to the next generation. (**B** and **C)** Fold-change of yellow (dashed lines) and green (solid lines) fluorescence intensity relative to ancestral YFP in each generation of phase I (**B**) and II (**C**). Error bars represent 1 SEM, from four replicate populations (single small symbols). **P* <0.05; ***P* <0.01; ****P* <0.001 [one-sided Dunnett tests with single-step adjustment to compare *S* with *W* (blue) or *N* (black)].

Fig. 2. Most rapid fixation of neofunctionalizing mutations after strong selection on the ancestral phenotype. (**A** and **B)** Emission spectra at the new excitation wavelength (405 nm) of mutants introduced into ancestral YFP (**A)** and U (**B).** The vertical axes indicate the fold-change of green fluorescence intensity relative to ancestral YFP at a given emission wavelength (horizontal axes). Each curved line represents the emission spectrum of 1 of the 44 mutants (n = 3). Colored lines in panel **A** indicate those emission-shifted mutations, and colored lines in panel **B** indicate mutations that significantly improved green fluorescence in the genetic background U (two-sided Dunnett test with single-step adjustment, *P*<0.05; fig. S5). Yellow and green vertical dashed lines indicate 530 nm (λex = 485 nm) and 512 nm (λex = 405 nm). (**C)** Frequencies of neofunctionalizing mutations G66S and Y204C or the double mutant U during evolution. Error bars represent 1 SEM from four replicate populations. **P* <0.05; ***P* <0.01; ****P* <0.001 [one-sided Dunnett tests with single-step adjustment to compare *S* with *W* (blue) or *N* (black)]. Abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; E, Glu; F, Phe; G, Gly; I, Ile; K, Lys; L, Leu; M, Met; S, Ser; V, Val; and Y, Tyr.

Fig. 3. Strong selection leads to greater mutational robustness and higher foldability than weak selection. (**A**) Yellow fluorescence retained by each population at the end of phase I after mutagenesis relative to its yellow fluorescence without mutagenesis. (**B**) Frequencies of cells fluorescing above background in yellow (left) (*17*) or above ancestral YFP (Anc) in green (right) (*17*) from each population at the end of phase I after mutagenesis. (**C** and **D**) Amount of soluble protein relative to ancestral YFP (**C**), and insoluble protein fraction (**D**) in each population at the end of phase I. (**E**) Recovery of yellow fluorescence over time (horizontal axis) during refolding of unfolded fluorescent proteins at 25 °C (*17*). Error bars represent 1 SEM based on four replicate populations. **P* <0.05; ***P* <0.01; ****P* <0.001 (two-sided Dunnett tests with single-step adjustment to compare *S* with *W*, *N* or ancestral YFP).

Fig. 4. Foldability-improving mutations enhance mutational robustness. (**A**) Recovery of yellow fluorescence during refolding of unfolded ancestral YFP and of the indicated variants at 25 °C (*17*). **(B)** Fluorescence intensity retained relative to fluorescence without mutagenesis (*17*). (**C**) Frequencies of fluorescence-positive cells after random mutagenesis (*17*). Error bars represent 1 SD (n=3). **P* <0.05; ***P* <0.01; ****P* <0.001 [two-sided Dunnett tests with single-step adjustment to compare each variant with ancestral YFP in **(B)** and **(C)]**.

Fig. 5. Enhanced mutational robustness promotes sweeps of adaptive mutations. (A) Percentage of cells with higher green fluorescence than ancestral YFP in populations *S*, *W*, and *N* after mutagenesis during each generation of evolution (*17*). **(B)** Increased mutational robustness (purple) rather than fluorescence increase (orange) is the major contributor to the acceleration of green fluorescence evolution by foldability-increasing mutations. **(C)** Strong selection purges nonsense mutations most efficiently during phase I evolution. Error bars represent 1 SEM from four replicate populations (**A** and **C**), or 1 SD over three biological replicates (**B**). **P* <0.05; ***P* <0.01; ****P* <0.001 [one-sided Dunnett tests with single-step adjustment to compare *S* with *W* (blue) or *N* (black)].

Fig. 6. Mutational robustness helps increase genetic diversity and evolvability. (A) Increase in the average number of amino acid changing mutations per fluorescent protein molecule relative to the beginning of phase II for evolving populations *S*, *W* and *N*. **(B)** Increase in genetic diversity relative to the beginning of phase II for evolving populations *S*, *W* and *N*. The data show the increase in the number of amino acid changing mutations (**A**) or in genetic diversity (**B**) in populations *S*, *W* and *N* at the end of phase II, relative to the end of phase I. Error bars represent 1 SEM from four replicate populations. **P* <0.05; ***P* <0.01; ****P* <0.001 [one-sided Dunnett tests with single-step adjustment to compare *S* with *W* (blue) or *N* (black)]. **(C)** Robustness-enhancing mutations can help evolving populations by-pass fitness valleys in an adaptive landscape. Robustness-enhancing mutations (short red arrow) can help move a population to a region of an adaptive landscape with low curvature, from where an adaptive peak can be reached more easily (long red arrow) than without robustness-enhancing mutations (blue arrow). Strong selection can help increase the frequency of robustness-enhancing mutations, and more so if such mutations also enhance fitness.

Supplemental materials:

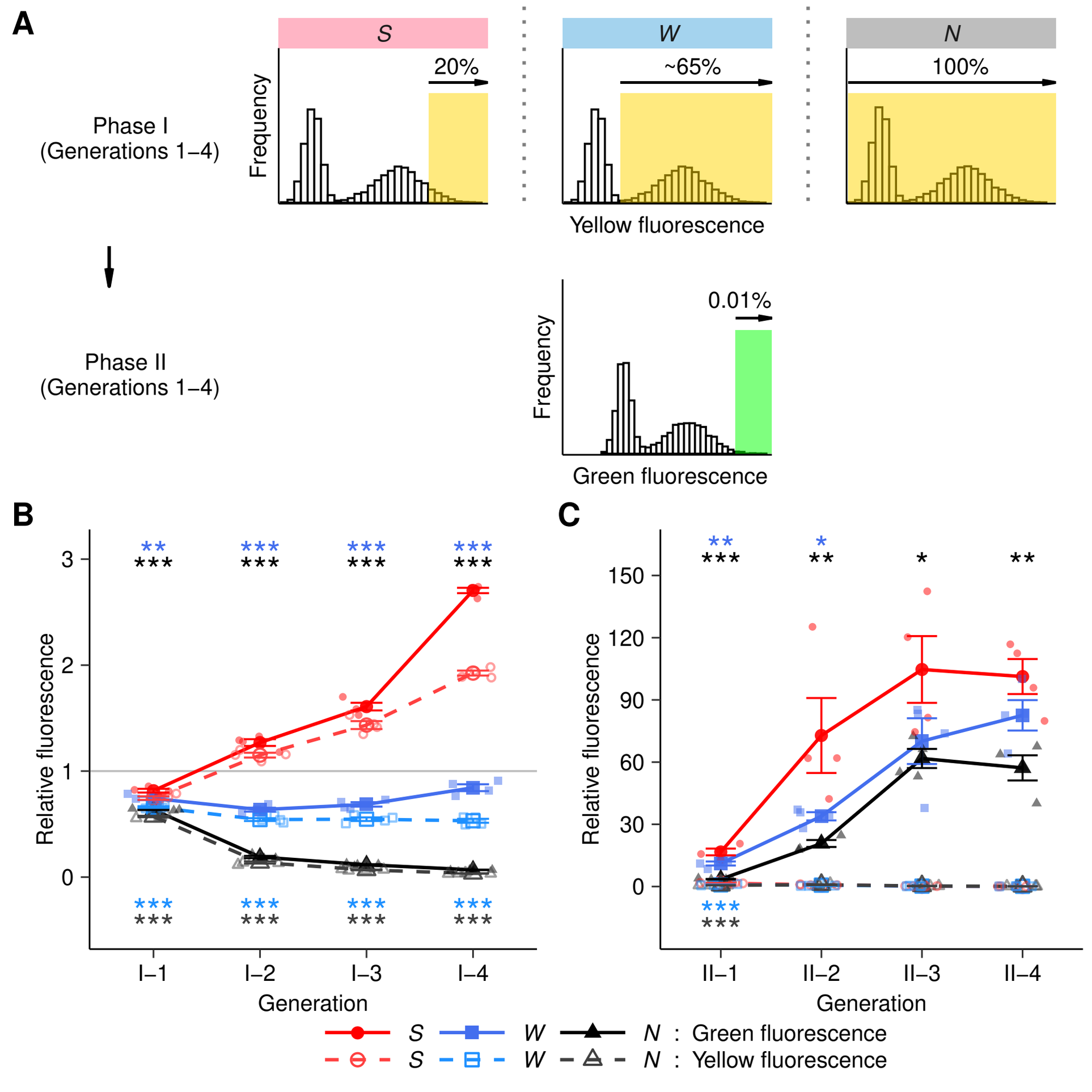
Materials and Methods

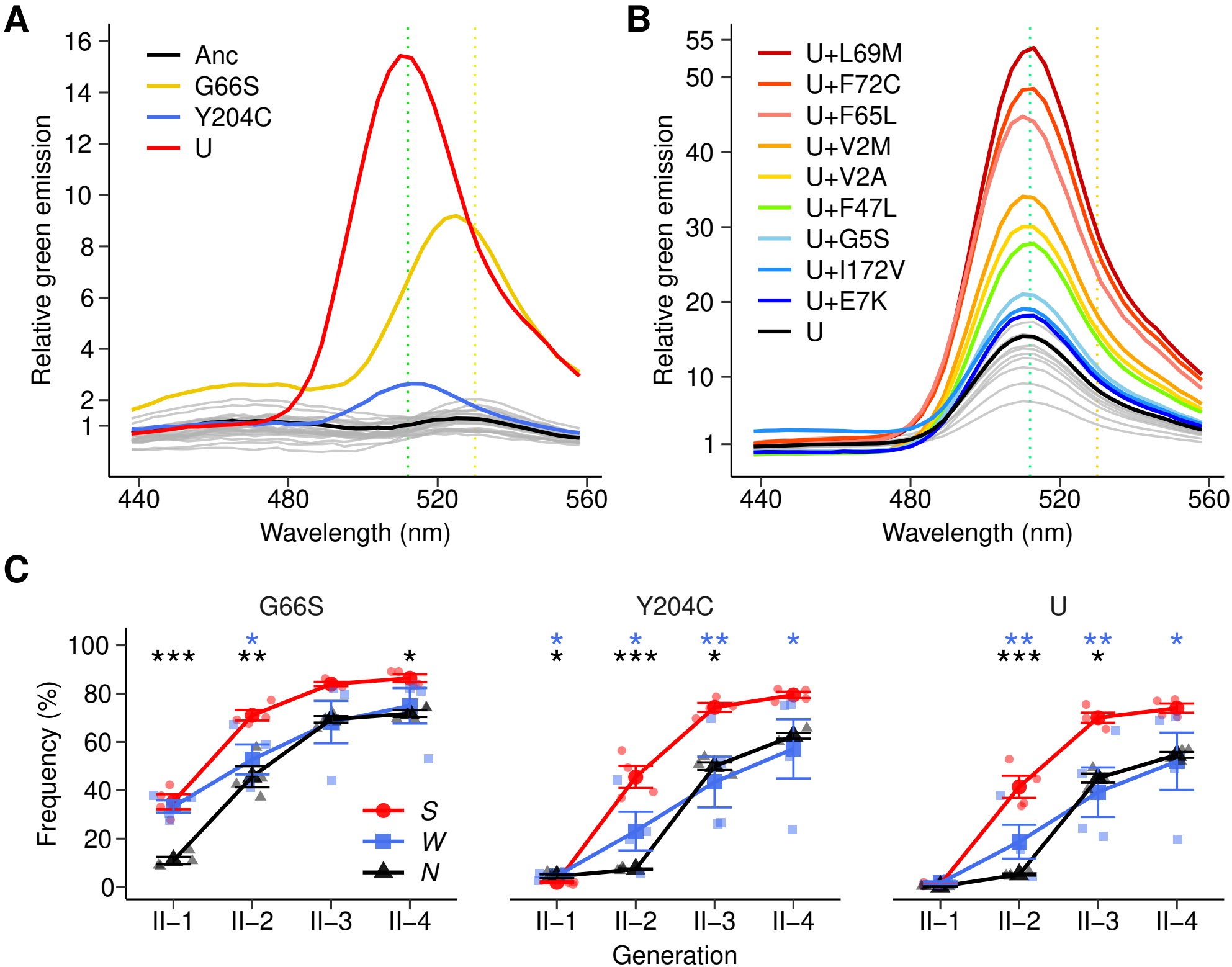
Figs. S1 to S21

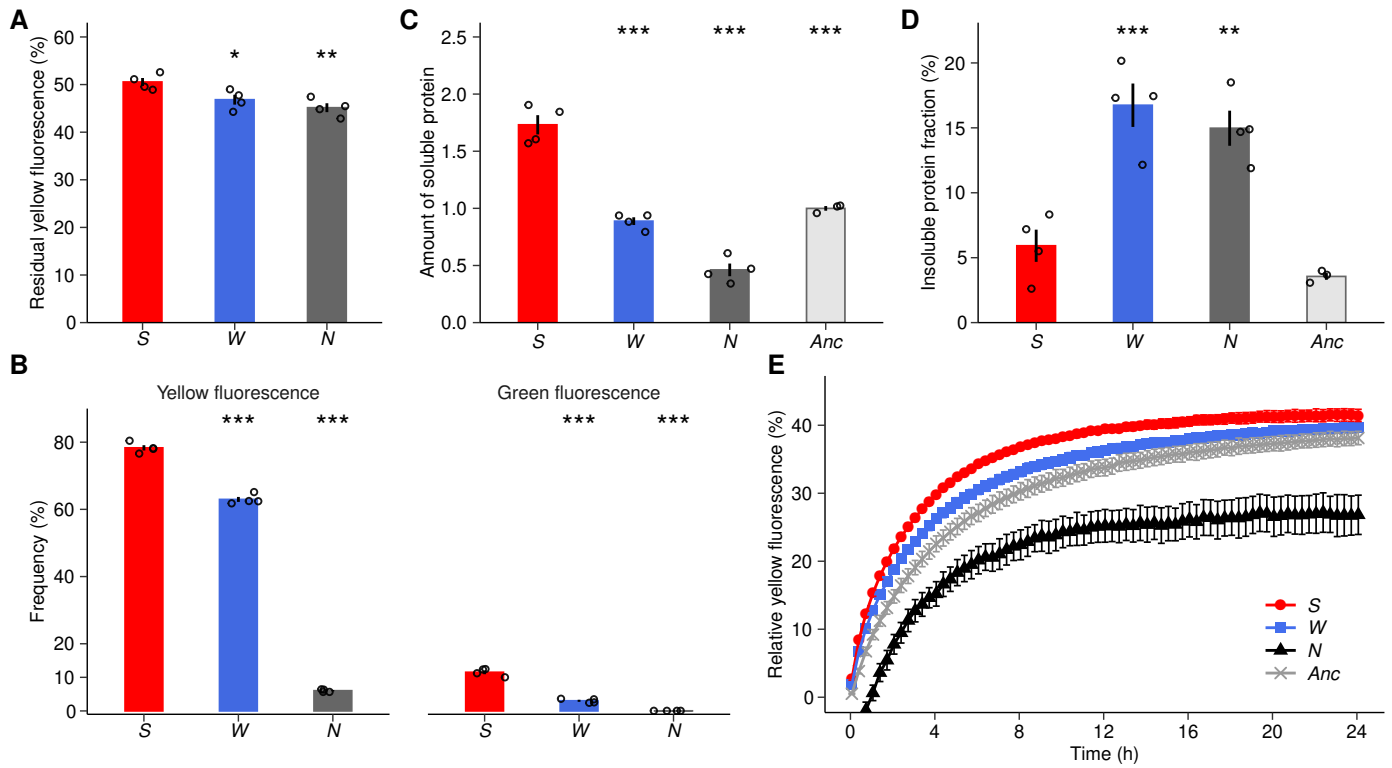
Tables S1 to S12

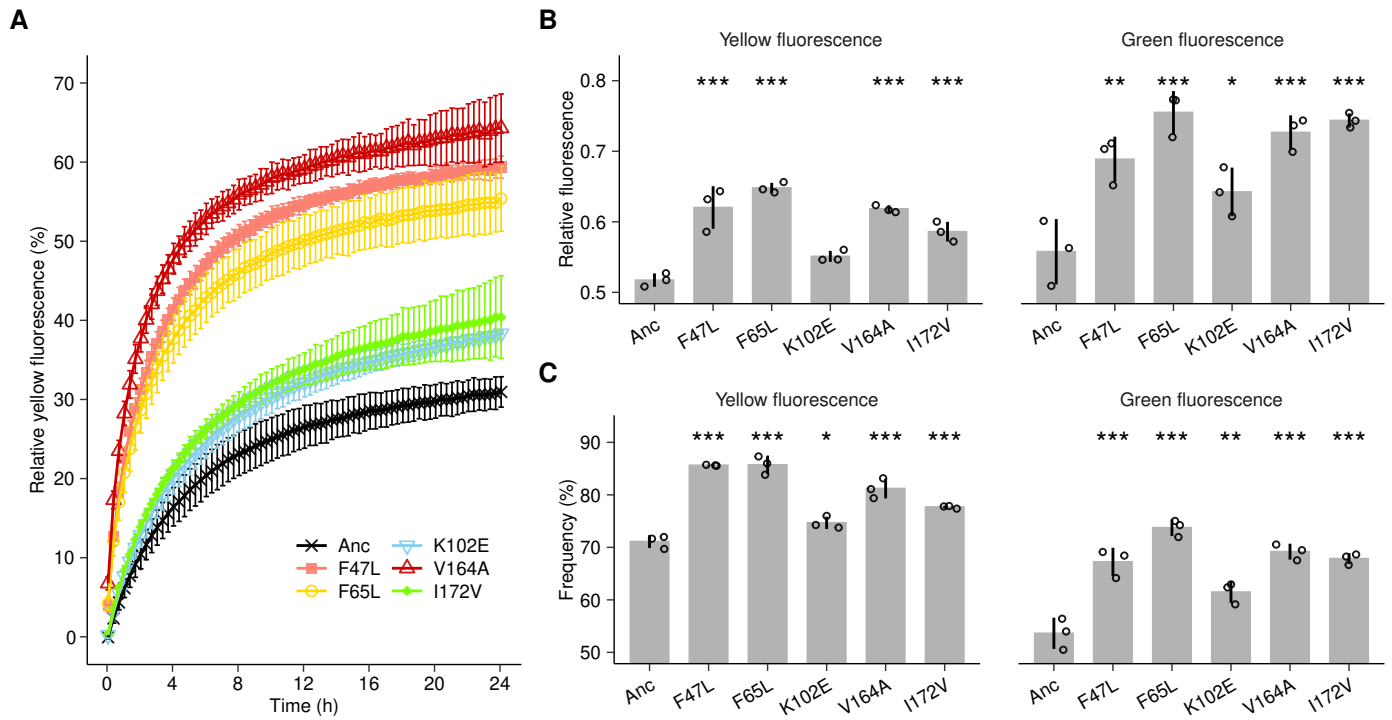
References (52-62)

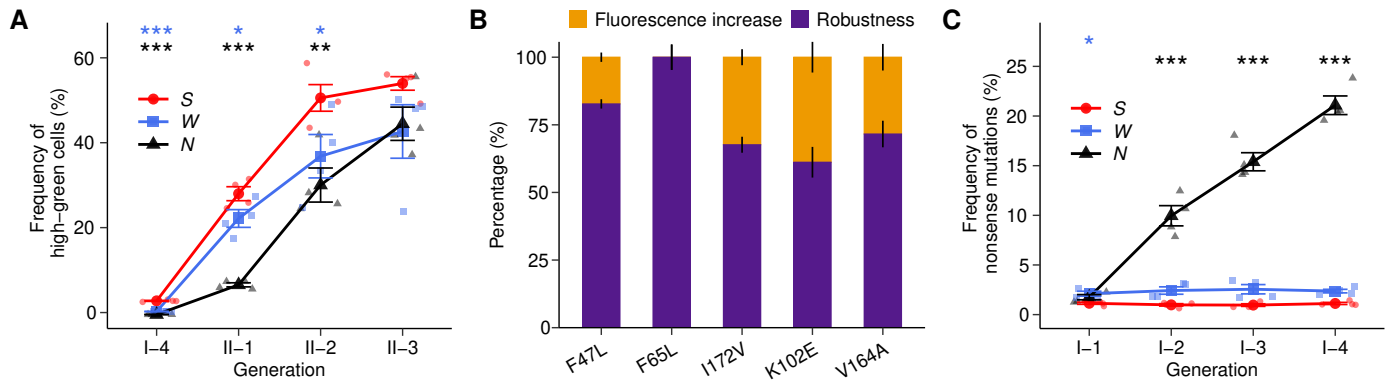
Database S1

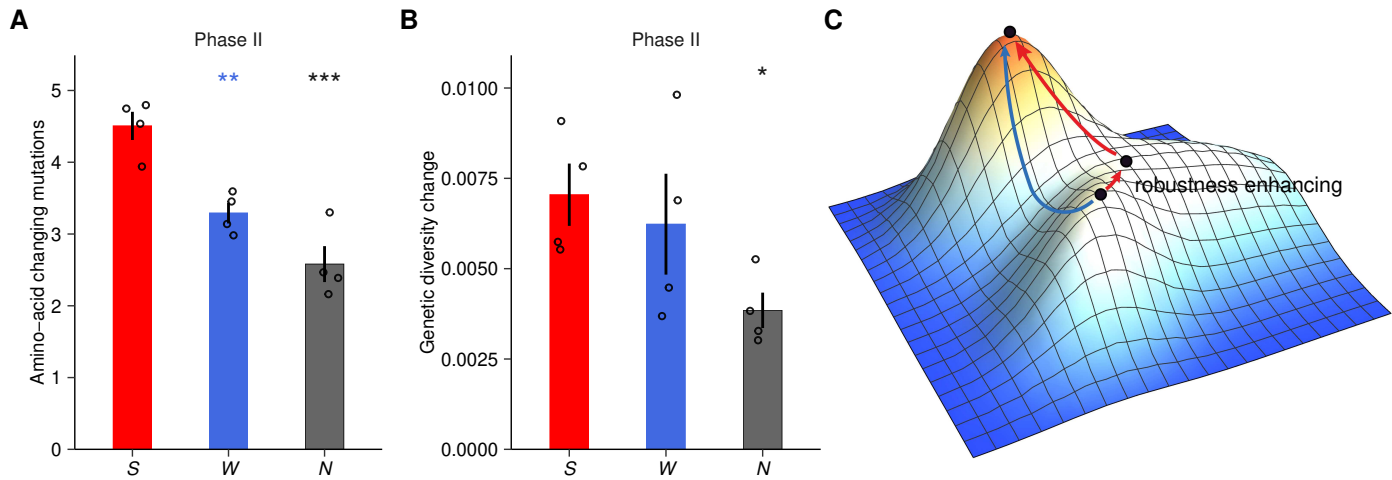

A
S
W
N
20%
~65%
100%
Frequency
Yellow fluorescence
Phase I
(Generations 1–4)
Phase II
(Generations 1–4)
0.01%
Frequency
Green fluorescence
B
C
Relative fluorescence
Generation
I–1
I–2
I–3
I–4
II–1
II–2
II–3
II–4
S
W
N : Green fluorescence
S
W
N : Yellow fluorescence


A
Anc
G66S
Y204C
U
Relative green emission
Wavelength (nm)
B
U+L69M
U+F72C
U+F65L
U+V2M
U+V2A
U+F47L
U+G5S
U+I172V
U+E7K
U
Relative green emission
Wavelength (nm)
C
G66S
Y204C
U
Frequency (%)
S
W
N
II–1
II–2
II–3
II–4
Generation


A
Residual yellow fluorescence (%)
60
50
40
30
20
10
0
*
**
S
W
N
C
Amount of soluble protein
2.5
2.0
1.5
1.0
0.5
0.0
***
***
***
S
W
N
Anc
D
Insoluble protein fraction (%)
20
15
10
5
0
***
**
S
W
N
Anc
B
Yellow fluorescence
Green fluorescence
Frequency (%)
80
60
40
20
10
0
***
***
***
***
S
W
N
E
Relative yellow fluorescence (%)
40
30
20
10
0
0
4
8
12
16
20
24
Time (h)
S
W
N
Anc


A
Relative yellow fluorescence (%)
Time (h)
Anc
F47L
F65L
K102E
V164A
I172V
B
Yellow fluorescence
Green fluorescence
Relative fluorescence
C
Frequency (%)


A
***
***
*
***
*
**
S
W
N
Frequency of high−green cells (%)
60
40
20
0
I−4
II−1
II−2
II−3
Generation
B
Fluorescence increase
Robustness
Percentage (%)
100
75
50
25
0
F47L
F65L
I172V
K102E
V164A
C
*
***
***
***
S
W
N
Frequency of nonsense mutations (%)
25
20
15
10
5
0
I−1
I−2
I−3
I−4
Generation


A
Phase II
**
***
Amino−acid changing mutations
0
1
2
3
4
5
S
W
N
B
Phase II
*
Genetic diversity change
0.0000
0.0025
0.0050
0.0075
0.0100
S
W
N
C
robustness enhancing